

Machine Learning basierte Response Style Identifikation: eine simulations-statistische Pilotstudie

Krause, Thomas

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Krause, T. (2020). *Machine Learning basierte Response Style Identifikation: eine simulations-statistische Pilotstudie*. (Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart -SISS-, 47). Stuttgart: Universität Stuttgart, Fak. 10 Wirtschafts- und Sozialwissenschaften, Institut für Sozialwissenschaften. <https://doi.org/10.18419/opus-10874>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

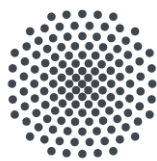
This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

SISS:
**Schriftenreihe des Instituts für Sozialwissenschaften
der Universität Stuttgart**

No. 47 / 2020

***Machine Learning basierte
Response Style Identifikation:
eine simulations-statistische
Pilotstudie***

Thomas Krause



Universität Stuttgart
**Abteilung für Soziologie
und empirische Sozialforschung**

SISS:
Schriftenreihe
des Instituts für Sozialwissenschaften
No. 47 / 2020

Machine Learning basierte Response Style Identifikation:
eine simulations-statistische Pilotstudie

Thomas Krause

ISSN 2199-7780

Universität Stuttgart
Institut für Sozialwissenschaften
Abt. f. Soziologie u. empirische Sozialforschung
70174 Stuttgart

Thomas Krause

**Machine Learning basierte Response Style Identifikation:
eine simulations-statistische Pilotstudie.**

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 47/2020.
(<https://www.sowi.uni-stuttgart.de/dokumente/forschung/siss/SISS47.pdf>)

Z U S A M M E N F A S S U N G: Response Styles stellen eine Herausforderung für die empirische Surveyforschung dar. Antwortverhalten, welches nicht mit dem Inhalt der Frage assoziiert ist, kann nicht nur Anteils- und Durchschnittswerte beeinflussen, sondern auch modellbasierte Parameterschätzung verzerren. In dieser Pilotstudie soll der Frage nachgegangen werden, ob sich über Machine Learning Verfahren ein empirisch basierter Ansatz zu einheitlichen Identifikation von den gängigsten Response Style Mustern konstruieren lässt. Dies soll aufwändige Kontrollverfahren, welche jeweils nur einzelne Muster finden können oder mit erheblichem Mehraufwand verbunden sind, ersetzen und somit eine universellere und praxistauglichere Option zu bisherigen Ansätzen darstellen. Der hier dargelegte Ansatz basiert auf der Kalibrierung des ML-Verfahrens anhand von synthetisierten Daten, welche der formalen Definition von Response Styles (RS) entsprechen und einem Anteil von empirischen Daten (European Social Survey), welche nicht von RS betroffen sind. Das hierauf trainierte Modell kann auf empirisch erhobene Daten angewendet werden, um RS-Muster in Survey-Daten zuverlässig entdecken und bearbeiten zu können. Die Ergebnisse der Studie legen mit Fehlerklassifikationsraten von 0.3 bis 3.5 % den ML-Ansatz als eine vielversprechende Alternative zu bisherigen Verfahren nahe.

Schlagworte: Response Style; Response Set; Machine Learning; Random Forest; Survey.

Thomas Krause

**Machine Learning based response style identification:
a simulation-statistical pilot study.**

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 47/2020.
(<https://www.sowi.uni-stuttgart.de/dokumente/forschung/siss/SISS47.pdf>)

A B S T R A C T: Response styles represent a challenge for empirical survey research. Response behavior which is not associated with the content of a survey question can not only influence proportional and average values but also distort model-based parameter estimation. In this pilot study we investigate whether machine learning (ML) methods can be used to construct an empirically based approach to a uniform identification of the most common response style patterns. This method could replace more complex control procedures, which can only find some patterns or are associated with considerable additional effort. Our approach could, therefore, be a more universal and practical option. The method presented here is based on the calibration of the ML model using synthesized data that correspond to a formal definition of response styles (RS) and a proportion of empirical data (European Social Survey) which is not affected by RS. The trained model can then be applied to empirically collected data in order to reliably detect and process RS patterns in survey data. With error classification rates of 0.3 to 3.5 %, the results of this study suggest that our ML-based response style detection is a promising alternative to previous methods.

Keywords: Response Style; Response Set; Machine Learning; Random Forest; Survey.

1 Einleitung

Machine Learning (ML) Verfahren werden inzwischen breit in der Praxis zur Betrugserkennung (fraud detection) eingesetzt. Darunter finden sich Anwendungsfälle im Bankenwesen, in der Versicherungsbranche, in Behörden und im Gesundheitswesen (vgl. zur Übersicht: Chandola et al., 2009: 13 ff.). Response Styles können konzeptionell als eine Form von Betrugsverhalten im Bereich der Surveyforschung verstanden werden. Statt ihre tatsächliche Einschätzungen oder Meinungen zu einem Frageitem zu äußern, werden von Befragten Antwortmuster gegeben, welche mit dem eigentlichen Inhalt des Surveys bzw. der Frage nichts zu tun haben. Während Interviewfälschungen – ebenfalls eine Form von Survey-Betrugsverhalten (data fabrication) – bereits mittels ML-Verfahren untersucht wurden (vgl. Birnbaum, 2012), finden sich zu Response Styles bisher noch keine Versuche der Bearbeitung mittels ML-Verfahren.

Diese Pilotstudie geht der Frage nach, ob sich über Machine Learning Verfahren ein ausschließlich empirisch basierter Ansatz zu Identifikation von den gängigsten Response Style Mustern konstruieren lässt. Dies könnte aufwändige Kontrollverfahren, welche jeweils nur einzelne Muster finden können oder zusätzliche Frageitems benötigen, ersetzen und somit eine universellere und praxistauglichere Möglichkeit zu bisherigen Ansätzen darstellen. Der hier diskutierte Ansatz basiert auf der Kalibrierung des ML-Verfahrens (bzw. Training) anhand von synthetisierten Daten, welche einer verteilungsbasierten Definition von Response Styles (RS) entsprechen und einem Anteil von empirischen Daten (European Social Survey), welche nicht von RS betroffen sind. Das hierauf trainierte Modell soll auf empirisch erhobene Daten angewendet werden können, um RS-Muster in Survey-Daten zuverlässig entdecken und bearbeiten zu können. In dieser Arbeit soll folglich im Kern der Frage nachgegangen werden, wie erfolgreich ML-Verfahren bei der Aufdeckung von RS-Antwortmustern sein können.

Auch wenn es divergierende Ansichten zum Ausmaß der Verzerrung von RS gibt, wäre es sicher wünschenswert, Fälle mit entsprechenden Antwortmustern, sofern möglich, zu identifizieren und entsprechend zu behandeln (bzw. auszuschließen). Darüber hinaus ermöglicht eine weiterführende Analyse von leistungsfähigen Identifikationsmodellen, Determinanten zu identifizieren, welche RS verursachen. Dies kann die Surveypraxis anleiten und zu entsprechenden Gegenmaßnahmen im Design der Untersuchung veranlassen, was zu zuverlässigeren Daten bei gleicher Erhebungsfallzahl führen kann. Zu diesem Zweck wird zunächst die Konzeption von RS in sozialwissenschaftlichen Daten diskutiert, sowie ursächliche Quellen von RS, Typisie-

rungsversuche und bisherige Identifikationsstrategien dargestellt. Im Anschluss daran werden die Grundprinzipien von sogenannten Supervised Machine Learning überblicksartig erläutert, um im nächsten Schritt die Anwendung auf Response Style Daten zu verdeutlichen. Nach dem Forschungsdesign und der Darstellung der Ergebnisse folgt die Diskussion mit einem Blick auf zukünftige Anwendungsmöglichkeiten für die Forschungspraxis.

2 Response Styles in sozialwissenschaftlichen Daten

Response Styles werden in der Regel von sogenannten Response Sets unterschieden. Dennoch werden die Begriffe in der Forschungspraxis teilweise als Synonyme behandelt (vgl. van Herk et al., 2004: 347). Wir bevorzugen im Weiteren eine Trennung der beiden Begrifflichkeiten, welche auf Rorer (1965) zurückgeht. Response Sets werden somit als Verzerrungsmuster verstanden, welche einen systematischen Zusammenhang mit dem Inhalt der Frage haben (z. B. Verzerrungen durch Soziale Erwünschtheit) (vgl. Schimmack et al., 2002: 463). Response Styles werden davon abgegrenzt, da ihr (Verzerrungs-)Muster nicht mit dem Inhalt von Fragetexten assoziiert ist und somit unabhängig vom eigentlichen Inhalt der Frage auftritt (vgl. auch Paulhus, 1990). Im Kontext einer Messung tragen sie somit zum systematischen Fehler bei (vgl. Jandura, 2018: 210). Längsschnittliche Analysen legen zudem den Schluss nahe, dass dieser systematische Fehler auch eine gewisse zeitliche Stabilität aufweist und somit entweder als ein stabiles individuelles Charakteristikum (vgl. z. B. Weijters et al., 2010b: 104) oder zumindest als konstantes Verzerrungsmuster verstanden werden kann. Neben der zeitlichen Stabilität ist gut dokumentiert, dass RS auch innerhalb eines Fragebogeninstruments stabil sind (vgl. Plieninger, 2017: 33). Eine alternative Perspektive auf Response Styles versteht diese als kommunikative Filter oder spezifische Formen von Kommunikationsstilen statt als Messfehler. Dabei wird teilweise argumentiert, dass sie als Ausdruck eines substantiell inhaltlich relevanten Faktors der General Response Style (GRS) verstanden werden können (vgl. He et al., 2014). Unabhängig von der inhaltlichen Deutung und Bedeutung von RS, ist eine möglichst zuverlässige Aufdeckung von RS im Rahmen einer empirischen Erhebung erstrebenswert. Zuvor werden zunächst Kategorisierungen und Determinanten von Response Styles überblicksartig dargestellt, bevor Maßnahmen und Strategien zu Identifikation diskutiert werden.

Typen, Verzerrungen und Quellen von Response Styles

Kategorisierungsversuche von Response Styles umfassen in der Regel eine Spanne von Mustertypen, welche mal mehr oder mal weniger praxisnahe und/oder überzeugend wirken. Zu Letzterem kann man wohl exotischere Varianten wie die (NARS) Net Acquiescence Response Style (die Tendenz mehr Zustimmung als Ablehnung zu zeigen) oder die (RR) Response Range (die Tendenz entweder eine weite oder eine enge Spanne an Antwortkategorien, um den Mittelwert zu verwenden) zählen. Eine breitere Beachtung in der Forschung haben eingängigere Typen von RS wie der (ERS) Extreme Response Style (Tendenz zur Wahl der Extremantwortkategorien), der (ARS) Acquiescence Response Style (Tendenz zur Zustimmungskategorien) oder der (MRS) Midpoint Response Style (Tendenz zur Mittelkategorie) bekommen (vgl. van Vaerenbergh & Thomas, 2013: 198). Daher werden in den weiteren Darstellung und Analysen auch vorrangig diese drei RS-Schemata diskutiert und analysiert.

Die Konsequenzen von RS umfassen Verzerrungen von Durchschnittswerten, der Varianz der Messungen und der multivariaten Zusammenhangsstruktur von mehreren Messungen. Während Antwortmuster nach ARS und ERS die Mittelwerte nach oben und unten verzerren, lenken MRS Muster Mittelwerte im Durchschnitt mehr in Richtung der Mittelkategorie. Diese Struktur setzt sich auch für die Varianz von Messungen fort. MRS reduziert hierbei die Varianz, wohingegen ERS und ARS sowohl eine Unter- wie auch eine Überschätzung der Varianz begünstigen können. Für multivariate Assoziationsstrukturen bei multiplen Messungen gilt, dass diese von MRS in der Regel (bzw. wenn mehrere Items von den Antwortmustern betroffen sind) verstärkt werden. Auch ERS und ARS werden die Zusammenhänge in der Regel verstärken, wobei dies zusätzlich von der Richtung des Itemwordings (insbesondere bei ARS) und der Konstanz der Extremkategoriewahl (bei ERS) abhängt (vgl. auch : Baumgartner & Steenkamp, 2001: 146 f.; van Vaerenbergh & Thomas, 2013: 197).

Mögliche Ursachenkomplexe für RS und die dahinterstehenden Mechanismen wurden in der relevanten Forschungsliteratur breit diskutiert und untersucht (vgl. hierzu im Weiteren auch van Vaerenbergh & Thomas, 2013: 198 ff.). An dieser Stelle kann und soll keine umfassende Darstellung hierzu gegeben werden. Es wird lediglich ein allgemeiner Überblick mit einem Verweis auf die relevante Literatur angestrebt.

Für eine grobe Einteilung der relevanten Determinanten scheint eine Unterteilung in erhebungsbezogene Größen und persönliche Faktoren angebracht und wird in der Regel anhand gleicher oder ähnlicher Kategorien vollzogen (vgl. z. B. : Baumgartner & Steenkamp, 2001: 144; Weijters, 2006: 29 ff.; van Vaerenbergh & Thomas, 2013: 198). Unter den erhebungsbe-

zogenen Größen finden sich situationelle Aspekte wie Zeitdruck bei der Befragung, Effekte erfahrener Interviewer oder die Form des Erhebungsverfahrens (z. B. Face-to-Face oder Web-survey), welche das Zurückgreifen auf verschiedene Formen von RS begünstigen. So zeigen sich Hinweise darauf, dass der Grad der Anonymität bei der Befragung die Tendenz falsche oder fiktive Antworten zu geben steigert, wobei dies bei Onlinebefragungen ein besonderes Problem darstellen kann (vgl. Jandura, 2018: 211 f.). Daneben werden aber auch formale Einflussfaktoren diskutiert wie das Skalenformat, Sprache des Erhebungsinstruments, das Labeling der Antwortkategorien (vgl. Weijters et al., 2010a: 244) oder Fragekomplexität, welche den «cognitive load» der Befragten beeinflussen, weshalb sie in Folge dessen eher auf RS zurückgreifen.

Unter den persönlichen Faktoren finden sich hingegen Prädiktoren von RS, welche dem Befragten selbst – oder zumindest dem gesellschaftlichen Kontext, in welchem der Befragte eingebettet ist – zugeordnet werden können. Darunter fallen z. B. themenspezifische Komponenten wie die persönliche Bedeutsamkeit des Themas für den Befragten, welche zu mehr ERS führen können (vgl. Gibbons et al., 1999: 377 f.), aber auch soziodemographische Aspekte. Für diese soziodemographischen Größen lässt sich ein zwiespältiges Bild bzgl. der Assoziation mit RS nachweisen. Für das formale Bildungsniveau zeigt sich zwar recht einheitlich, dass höhere Bildung (ggf. als Proxy für komplexes Denken oder Differenzierungsfähigkeit) zu einem geringeren Niveau an RS führt (vgl. z. B. van Vaerenbergh & Thomas, 2013; Meisenberg & Williams, 2008: 1547 ff.), aber für andere Determinanten sieht das Bild deutlich unklarer aus. So fallen die gefundenen Effekte bzgl. des Alters und des Geschlechts des Befragten äußerst gemischt aus (vgl. Rammstedt et al., 2017: 194; He et al., 2014: 315; van Vaerenbergh & Thomas, 2013: 202 f.), so dass nach jetzigem Stand nicht klar gesagt werden kann, ob das Alter oder Geschlecht zu mehr RS führen. Ganz allgemein scheinen persönliche Kenngrößen nur einen geringen Anteil am Auftreten von RS (statistisch) erklären zu können. Die Schätzungen bewegen sich im Rahmen von 1 bis 8 % Varianzaufklärung (vgl. van Vaerenbergh & Thomas, 2013: 205).

Neben den individuellen Aspekten scheint der soziale bzw. nationale Kontext eine besondere Relevanz zu haben. Der aktuelle Konsens der Forschung ist, dass RS mit kulturellen Größen deutlich assoziiert sind und traditionelle Kulturkreise zu mehr RS neigen (vgl. Rammstedt et al., 2017: 191). van Herk et al., 2004 können beispielsweise für Europa zeigen, dass RS in südlichen Ländern präsenter sind als in nördlichen Ländern. Meisenberg & Williams, 2008 zeigen das RS in weniger entwickelten Ländern am weitesten verbreitet sind und dass

das ARS-Muster am besten durch die Korruptionsrate des Landes (statistisch) erklärt werden kann. Rammstedt et al. (2017) können hingegen anhand von Daten des European Social Survey belegen, dass der Grad an Kollektivismus und Korruption bis zu 74 % der Länderunterschiede bzgl. RS erklären können.

Die hier dargestellte Assoziationsstruktur zwischen Response Styles und den jeweils betrachten (potenziellen) Ursachen basiert im Kern auf der möglichst akkuraten Messung bzw. Identifikation von Response Styles. Welche bisherigen Verfahren dafür eingesetzt werden, soll im Folgenden diskutiert werden.

3 Bisherige Ansätze und Probleme bzgl. der Identifikation von Response Styles

Ansätze zur Identifikation von Response Styles lassen sich entlang der ihnen zugrundeliegenden Informationsbasis differenzieren. Nach unserer Auffassung können die Verfahren dabei entweder zu den Zusatz-Item-Ansätzen oder zu den Latenten-Struktur-Ansätzen gezählt werden.

Die Informationsbasis von Zusatz-Item-Ansätzen gründet auf der Berücksichtigung von zusätzlichen Frageitems, welche keine inhaltliche Relevanz für die Fragestellungen von Surveys haben, sondern lediglich auf die Identifikation von RS ausgerichtet sind. Dazu zählen beispielsweise:

- Zählung doppelter Zustimmung bei gedrehten Items. Nach diesem rudimentären Kalkül werden zustimmende (analog ablehnende) Bewertungen auf einer Ratingskala für Frageitems, welche die gleiche inhaltliche Dimension/Größe berühren aber hinsichtlich des Frage-textes einmal positiv und einmal negativ formuliert wurden, als ein Maß für ARS verwendet (vgl. z. B. Johnson et al., 2005: 270). Bei diesem Vorgehen müssen Fragen zu gleichen Sachverhalten mit einer positiven und negativen Itemformulierung bei der Erhebung vorhanden sein. Folglich ist das Problem von negativ formulierten Frageitems, dass man sie noch vor der eigentlichen Datensammlung in das Erhebungsinstrument einbauen muss und dies für Sekundärdaten oft nicht möglich ist. Auch bei Primärerhebungen muss beachtet werden, dass zusätzliche Items auch i. d. R. mit zusätzlichen Kosten verbunden sind. Des Weiteren bringt die negative Formulierung die Gefahr mit sich, dass die Frageitems von einem Teil der Befragten fehlerhaft interpretiert werden, was die Messung an und für sich potentiell

kompromittiert und dies unabhängig davon ob RS vorhanden sind oder nicht.

- Die Methode der «Representative Indicators of Response Styles» (RIRS) nutzt eine Reihe von zusätzlichen Frageitems, welche inhaltlich maximal heterogen sind. Die Grundidee dieses Ansatzes ist es, den inhaltlich substantiellen Zusammenhang der Items zu minimieren. Die inhaltliche Kovarianz zwischen den Items sollte dabei möglichst bei Null liegen (abgesehen von zufälligen Assoziationen). Wenn alle Elemente verschiedene Konstrukte repräsentieren, welche nicht miteinander in Beziehung stehen, kann man annehmen, dass es keine andere Konsistenz/Korrelation in den Antworten existiert als diejenige, welche durch Response Styles hervorgerufen wird (vgl. Weijters, 2006: 42). Die empirische Zusammenhangsstruktur und die Antwortmuster können dann als Maß für verschiedene RS-Typen herangezogen werden. Der Nachteil dieses Verfahrens ist, dass viele zusätzliche (und zudem unkorrelierte) Frageitems benötigt werden. Dies führt sowohl bei der Fragebogenkonstruktion als auch bei der letztendlichen Durchführung der Befragung zu Mehrkosten.
- RIRMACS basieren auf der gleichen Grundidee wie RIRS, verbinden diese allerdings noch mit der Modellierung von verschiedenen Typen von RS als latente Variablen im Rahmen eines Strukturgleichungsmodells. Dabei können die latenten RS-Konstrukte sowohl im Rahmen einer konfirmatorischen Faktoranalyse eingesetzt werden als auch als Kovariate im Rahmen einer weiteren strukturellen Analyse (vgl. Weijters et al., 2008).

Zusammenfassend kann konstatiert werden, dass Zusatz-Item-Ansätzen das offensichtliche aber weitreichende Problem haben, dass sie zusätzliche Items oder vorherige Berücksichtigung bei den Itemformulierungen benötigen. Damit sind eine Reihe von Nachteilen verbunden, welche je nach Art und Form der geplanten Analyse zu zusätzlichen Kosten und Fehlerquellen bis hin zu unüberwindbaren Hürden führen. Die Berücksichtigung von negativ (bzw. gedreht) formulierten Frageitems kann bei den Befragten zu Interpretationsproblemen hervorrufen und gleichzeitig zeitliche und finanzielle Mehrkosten der Befragung verursachen. Diese Mehrkosten sind für die RIRS und RIRMACS Methodik noch deutlich ausgeprägter, da eine Vielzahl von Items benötigt wird. Zusätzlich ist man mit dem Problem konfrontiert, dass diese maximal heterogenen Items erst einmal vom Forschenden zusammengestellt werden müssen, da sie nicht allgemeingültig für jeden Themenkomplex gelten. Darüber hinaus gilt es zusätzlich erhebungstechnische Aspekte abzuwägen, wie Ermüdungserscheinungen oder Dropoutwahrscheinlichkeiten der Befragten bei längeren Befragungen. Eine gänzliche Hürde für diese Zusatz-Item-Ansätze sind in der Regel Sekundäranalysen, für welche der Forschung post hoc keinen

Einfluss mehr auf die Frageitems hat und eine darauf basierende RS-Analyse unmöglich ist.

Die Informationsbasis für Latente-Struktur-Ansätze speist sich aus der Annahme von latenten Antwortmutterpopulationen und/oder latenten Methodeneinflussfaktoren, welche modellbasiert auf Grundlage von inhaltlich substantiell bedeutsamen Frageitems in Survey (bzw. deren Varianzstruktur und Antwortmuster) abgeschätzt werden. Die Identifikation beruht also nicht auf dem Vergleich von inhaltlich relevanten Items mit zusätzlichen Methodenitems, sondern lediglich auf der inhaltlich vorgegebenen Itemstruktur. Dazu zählen

- in der einfachsten Variante die Count-Methodik, welche bei fehlender Varianz der Antwortmuster einer Messeinheit entweder von ARS, ERS oder MRS ausgehen. Dabei werden über die Messeinheiten Summenscores gebildet, welche ab einem gewissen Schwellenwert nicht mehr als inhaltliche Antworten behandelt werden, sondern als RS-Muster.
- Elaborierte Varianten im Kontext von Strukturgleichungsmodellen (SEM) arbeiten mit latenten Methodenfaktoren, welche neben den inhaltlich substantiellen Faktoren, spezielle latente Methodenfaktoren mit beinhalten (vgl. z. B. Saris et al., 2004; Saris & Aalberts, 2003; Billiet & McClendon, 2000). Die Antwort auf ein manifestes Frageitem wird somit von inhaltlichen Unterschieden bestimmt als auch von Effekten von ARS (und tlw. ERS). Die Varianz der Items wird somit konzeptionell in inhaltlich bedeutsame Varianz, Response Style basierte Varianz und Restfehler Varianz aufgeteilt.¹Konzeptionell zwar etwas unterschiedlich, aber rechnerisch sehr ähnlich verfahren multivariate Item Response Theory (IRT) Modelle. Auch hier wird modellbasiert versucht, zwischen inhaltlich bedeutsamer Unterschiedlichkeit und RS-basierter Unterschiedlichkeit zu differenzieren (vgl. z. B. Falk & Cai, 2016; Plieninger & Meiser, 2014; Wetzel & Carstensen, 2017).
- Finite Mixture Modelle oder Latent Class Regression verstehen Response Styles nicht als eine kontinuierliche latente Variable wie im SEM- oder IRT-Paradigma, sondern folgen einem Ansatz, welcher von der Idee als eine Form von Clusteranalyse (allerdings ohne feste Klasseneinteilung) verstanden werden kann (vgl. z. B. Eid & Rauber, 2000; Meiser & Machunsky, 2008; Moors, 2010; Wetzel et al., 2013). Die Antwortmuster der manifesten Indikatoren verweisen konzeptionell auf diskrete «Subpopulationen» statt einer zugrunde liegende kontinuierlichen Größe. Anhand von verschiedenen Maßzahlen, welche i. d. R. auf

¹Im Rahmen eines SEM Faktormodells können ERS und ARS ebenfalls als Non-Invarianz der Faktorladungen und fehlende Invarianz der Intercepts verstanden werden. Dabei wird Äquivalenz der Messung durch den systematischen Einfluss auf das Messmodell konzeptualisiert.

der jeweiligen Modellpassung basieren, kann dann entschieden werden, ob den Daten weitere latente Subpopulationen zugrunde liegen, welche in diesem Anwendungsfall Personengruppen mit RS-Muster wären. Die durch das Mixturemodeling gewonnenen Wahrscheinlichkeiten für die latenten Klassenzugehörigkeiten können im nächsten Schritt zum Ausschluss oder Behandlung von RS-Fällen verwendet werden.²

Allerdings sind auch mit den Latente-Struktur-Ansätzen Probleme verbunden. Die einfache Count Technik hat beispielsweise vorrangig das Problem, dass alle Items einer Skala oder alle untersuchten Items ein entsprechendes RS-Muster aufweisen müssen, damit ein Muster identifiziert werden kann. Werden nur wenige Items nicht musterkonform beantwortet, kann ein entsprechender Fall nicht mehr gefunden werden. Folglich sind nur Extremfälle identifizierbar, was im Durchschnitt zu einer höheren Falsch-negativ-Rate führen kann. Die Sensitivität der Analyse wird damit eingeschränkt.

Für modellbasierte Strukturgleichungsansätze mit entsprechenden Methodenfaktoren ergeben sich Probleme technischer und praktischer Natur. So benötigt die Spezifikation von Methodenfaktoren im Rahmen von Strukturgleichungsmodellen Freiheitsgrade, welche nicht in jedem Analysesetting in ausreichenden Umfang vorhanden sein werden. Selbst wenn entsprechende Freiheitsgrade bei der Analyse genutzt werden können, steigert dies die Modellkomplexität, was wiederum im Schnitt die Modellkonvergenz der Gesamtschätzung schwieriger macht und damit die Untersuchung des inhaltlichen Parts des Modells gefährdet. Die praktischen Einschränkungen sind auch dafür verantwortlich, dass oft nur nach einer spezifischen Form von RS-Muster kontrolliert werden kann, wodurch andere RS-Formen unberücksichtigt bleiben müssen.

Finite Mixture Modelle oder Latent Class Regressionen teilen das Problem, dass die Grenzziehung zwischen den latenten Klassen von RS-Fällen und tatsächlich inhaltlichen Antwortmustern nicht eindeutig geschehen kann und sich nur bedingt standardisieren oder stichhaltig begründen lässt. So können neben methodischen Aspekten wie RS-Antwortmustern auch nicht entdeckte inhaltliche Subpopulationen für zusätzliche latente Klassen verantwortlich sein. Die Anzahl an Mixture-Klassen werden i. d. R. anhand von Kriterien bzgl. der Modellpassung sowie dehnbaren Leitlinien bestimmt, ohne dass eindeutige und universalgültige Kriterien existieren (vgl. z. B. Krause, 2019: 26 ff.). Darüber hinaus wird zudem häufig davon abgeraten, dass Klassen, welche weniger als 5% der Gesamtfallzahl beinhalten nicht berücksichtigt

²Der Vollständigkeit halber sei erwähnt, dass Mixture Models in manchen Klassifikationen auch als eine Form von Unsupervised Machine Learning verstanden wird.

werden sollten (vgl. Wickrama et al., 2016: 216 f.). Verbindet man diese Empfehlung mit Prävalenzschätzungen von RS-Antwortmustern unterhalb dieses Prozentwerts mit der Vielzahl an Entscheidungsheuristiken und Maßzahlen, wird deutlich, dass die Identifikation entsprechender Antworten mit viel Unsicherheit und Interpretationsspielraum verbunden ist, womit eine klare Trennung in inhaltlich relevante Antworten und Response Style bedingte Antworten oft nicht möglich sein wird.

4 ML-Verfahren für Response Styles

4.1 Prinzipien von ML-Verfahren

Bei Machine Learning Verfahren muss grundsätzlich zwischen supervised und unsupervised unterschieden werden (vgl. Ghani & Schierholz, 2017: 151 ff.). Während supervised Verfahren (sML) keine klare Zielgröße (auch: Label) haben, operieren unsupervised ML-Verfahren (uML) ohne eine vorher klar abgesteckte Zieldimension. uML-Verfahren stellen somit eine eigene Kategorie von strukturentdeckenden Verfahren dar, welche hier nicht diskutiert werden. Im Folgenden geht es demnach um die Unterschiede zwischen sML-Verfahren und klassischen sozialwissenschaftlichen statistischen Analysemodellen.

Supervised Machine Learning Verfahren folgen einem eigenen Paradigma, welches mit den etablierten Verfahren der Sozialforschung bisher großenteils nicht übereinstimmen. ML-Verfahren setzen den Fokus auf die Vorhersage der Zielgröße. Klassische statische Modelle aus der Sozialforschung (z. B. lineare, logistische Regressionen oder Strukturgleichungsmodelle) beinhalten zwar auch eine Vorhersagekomponente, setzen den Fokus aber auf die Signifikanztests der Effekte zwischen den unabhängigen Größen und dem abhängigen Zielkriterium. Bei klassischen Verfahren steht der Test von a priori festgelegten Hypothesen im Vordergrund, welche im Rahmen des statistischen Modells abgebildet werden und anhand von Signifikanztests auf deren Übertragbarkeit auf die zugrundeliegende Grundgesamtheit überprüft werden.³ Das primäre Ziel von sML-Verfahren stellt hingegen die Steigerung der Prognosegenauigkeit dar. Im Vorhinein festgelegte Hypothesen sind im allgemeinen nicht Teil der (oder Anwendungsfall von) ML-Modellen. Viel mehr wird das ML-Modell anhand der Daten generiert. Um bei dieser Suche nach Vorhersagegenauigkeit ein Overfitting des Modells auf die individuelle Stichprobe bzw. Datenlage mit ihren zufallsbedingten Idiosynkrasien zu vermeiden, findet in der Regel eine Bewertung der Modelle anhand einer Validierungsstichprobe statt, welche bei

³Mit einer angebbaren Irrtumswahrscheinlichkeit bei diesem Inferenzschluss.

der Generierung/Schätzung des ML-Modells nicht verwendet wurde. Diese Validierungsstichprobe kann entweder durch ein vorheriges manuelles Auftrennen der Daten in Trainings- und Validierungsstichprobe geschehen oder durch eine weitere Datenerhebung.

Die breiteste Anwendung von supervised ML-Verfahren finden aktuell Random Forest Modelle, Gradient Boosting Verfahren⁴ und neuronale Netzwerke. Während neuronale Netze insbesondere für unstrukturierte Datenstrukturen geeignet sind (z. B. Bilddaten, natürliche Sprachdaten, etc.) und hier gute Vorhersageergebnisse erzielen, sind Random Forest Modelle und die Gradient Boosting Trees besonders gut auf strukturierte Daten anwendbar, wie man sie beispielsweise in Datensätzen der klassischen quantitativen Sozialforschung (Querschnitts- und Panelerhebungen) findet. Durch ihr Verfahrensprinzip können ML-Verfahren Modelle generieren, welche die Zielgröße besser (oder bei perfekten klassischen Modellen zumindest gleich gut) prognostizieren können als klassische Modelle. Die Anpassung an die Struktur der Zusammenhänge zwischen Daten wird durch die ML-Verfahren iterativ perfektioniert, ohne dass eine spezifische Modellstruktur bekannt oder vorgegeben werden muss. Auch die Anzahl der zu berücksichtigenden Variablen stellt für ML-Verfahren keine Hürde dar. Daher eignen sie sich auch besonders für Daten-Situationen in welchen die Anzahl der Prädiktoren hoch und die Fallzahl gering ist (small N, big p). Klassische statistische Modelle der Sozialforschung geraten hierbei hingegen an ihre Grenzen.

Einer der zentralen Vorteile dieser ML-Verfahren ist, dass die Form der Assoziation zwischen den Prädiktoren und der Zielgröße nicht im Vorhinein festgelegt werden muss. Interaktionsbeziehung oder nicht-lineare Zusammenhänge werden durch die Verfahren berücksichtigt, ohne dass der Forschende selbige im Vorhinein spezifiziert oder die Verlaufsform des Zusammenhangs bestimmt. Ein wichtiger Nachteil von ML-Verfahren im selbigen Zusammenhang ist allerdings die Undurchsichtigkeit (Stichwort: Opazität von ML-Verfahren). Auch wenn diese Modelle nicht-lineare Zusammenhänge und Interaktionsbeziehungen adäquat bei der Modellgenerierung berücksichtigen können, gibt das Ergebnis der Modellschätzung in der Regel keinen einfach interpretierbaren Hinweis darauf, wie diese Beziehungen nun tatsächlich aussehen, wie man es von klassischen statistischen Modellen der Sozialforschung gewohnt ist. Besonders neuronale Netze sind dafür berüchtigt, dass der Mechanismus bzw. der Zusammenhang zwischen Input und Output nicht nachvollziehbar ist. Gleiches gilt allerdings auch für die tree-basierten Verfahren Random Forest und Gradient Boosting. Auch wenn ihre grundlegende Verfahrenslogik und eingesetzte Schätzmethodik klar und nachvollziehbar ist, trifft

⁴In der Regel auf Tree-Verfahren angewendet.

dies auf die durch sie individuell generierte Zusammenhangsstruktur nicht zu. So können alle generierten Modellparameter zwar ausgelesen werden, aber das Format ist in der Regel nicht «forscherlesbar».⁵ Der Fokus der Verfahren liegt damit, wie erwähnt, nicht auf der Dekonstruktion und Explikation der Zusammenhänge, sondern auf der korrekten Prognose der jeweiligen Zielgröße. Dieses Manko von ML-Verfahren gehört wohl mit zu den zentralen Gründen, wieso sie in der Sozialforschung nicht schon eine breitere Anwendung gefunden haben.

4.2 ML-Verfahren zur Identifikation von Response Styles

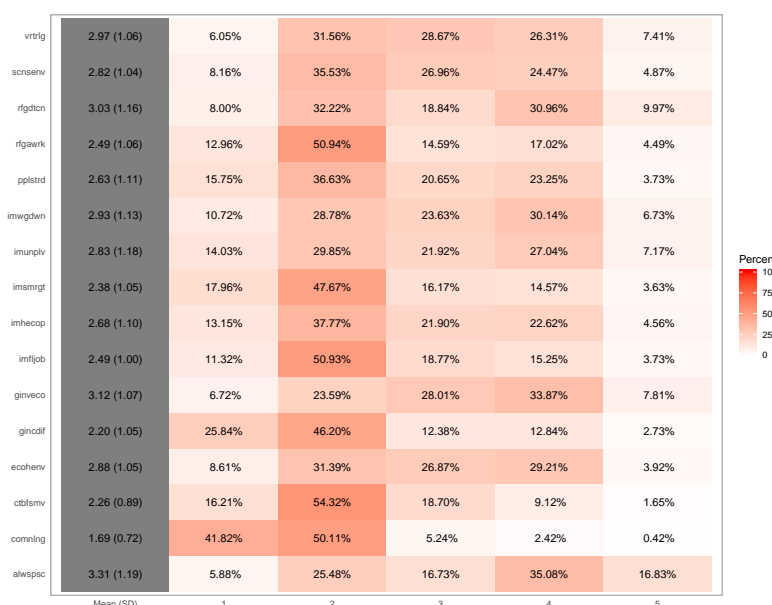
Für den Anwendungsfall der Response Style Entdeckung in Datensätzen haben ML-Verfahren den Vorteil, dass weder Annahmen zur Zusammenhangsstruktur von einzelnen Frageitems noch sonstige Modellierungs- oder Modellannahmen im Vorhinein festzulegen sind. Die Kategorisierung zwischen RS und regulären/inhaltlichen Antwortmustern erfolgt auf empirischer Basis. Wenn zum Training des Verfahrens ausreichend viele Beispielfälle für Response Styles und für Fälle, welche keinen Response Style enthalten, vorhanden sind, sollte dies ausreichen, um ein Modell anhand der Daten zu generieren. Dieses Modell kann anschließend auf empirisch erhobene Daten angewendet werden um den RS-Status (möglichst zuverlässig) zu prognostizieren und somit die Datenbereinigung oder den Umgang mit RS-Mustern anzuleiten. Da es sich bei Befragungsdaten um hoch strukturierte Daten mit vordefinierten Kategorienschemata handelt, bietet sich der Einsatz von Random Forest Verfahren an. Zudem zeigt die praktische Anwendung, dass mit Random Forest Modellen eine hohe Genauigkeit bei der Klassifikation erreicht werden kann ohne gleichzeitig ein übermässiges Gewicht auf (möglicherweise fehlerhafte) Ausreißerfälle zu legen (vgl. auch Ghani & Schierholz, 2017: 170).

Der hier verfolgte Design basiert auf der Kalibrierung des ML-Verfahrens (bzw. Training) anhand von simulierten Daten, welche der Definition von Response Styles (RS) entsprechen und einem Anteil von empirischen (nicht-simulierten) Daten, welche explizit nicht von RS betroffen sind. Das hierauf trainierte Modell soll anschließend auf die empirisch erhobenen Daten angewendet werden können um RS in den Daten zu entdecken. Dabei wird zwischen vier Simulations-Szenarien mit einem unterschiedlich hohen Anteil an von RS betroffenen Frageitems unterschieden.

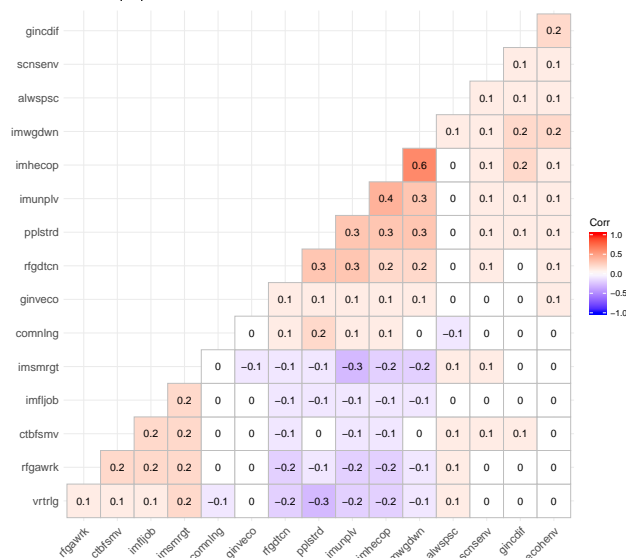
Die empirische Datengrundlage dieses Papers ist der European Social Survey 2002 (ESS

⁵Zumindest gilt dies, wenn die Verfahren ein gewisses Mindestmaß an Komplexität übersteigen (was bei der praktischen Anwendung immer der Fall sein wird).

2002). Diese Datenbasis wurde gewählt, da sie äußerst viele empirische Fälle umfasst, geeignete und ausreichend viele Skalenitems zur Betrachtung von Response Styles beinhaltet und bereits in vorangegangenen Arbeiten zu Response Styles untersucht wurde, womit Vergleichbarkeit mit vorausgehenden RS-Analysen hergestellt werden kann (vgl. Rammstedt et al., 2017). Zum Test und zur Modellkonstruktion wurden 16 Skalenitems aus dem ESS verwendet, welche bereits in vorangegangenen Studien ausgewählt wurden. Alle Skalenitems wurden anhand einer fünfstufigen Ratenskala bewertet. Die Items umfassen Fragekomplexe, welche die Bereiche Umwelt, Immigration, kulturelle Kohäsion und Wirtschaft berühren.⁶



(a) Rating-Distribution of ESS 2002



(b) Pairwise Missingness, N=(37778 bis 41919).

Abbildung 1: Ratingskalaverteilung und Korrelationsmatrix des ESS 2002

⁶Die Frageformulierungen und Variablenbezeichnungen können aus Tabelle 4 entnommen werden.

Aus Abbildung 1 kann die assoziative Struktur und die Verteilung auf der Ratingskala der 16 Frageitems aus der empirischen Datenbasis entnommen werden. Hieraus wird ersichtlich, dass Variablen enthalten sind, welche miteinander sowohl positiv wie auch negativ assoziiert sind, aber auch ein gewisser Grad an freier Varianz zwischen den Messungen besteht. Für die Generierung des ML-Models wurden die empirischen Daten um diejenigen Fälle reduziert, welche auf diesen Frageitems ein Muster aufzeigten, welches entweder dem RS-Muster ERS, ARS oder MRS entspricht (Anhand der einfachen Countmethodik). Die nach dieser Reduktion übrigen Fälle im Datensatz werden als Referenz für nicht vorhandene RS-Muster behandelt. Dabei ergibt sich jedoch das konzeptionelle Problem, dass die einfache Count-Methodik nicht sicher stellen kann, dass keinerlei von RS betroffene Daten in den reduzierten empirischen Daten verbleiben. Leichte Abweichungen von extremen Antwortmustern auf den 16 Frageitems nach ERS, ARS und MRS können zu einer Klassifikation als Non-RS führen, obwohl die restlichen Items von RS betroffen sind. Für Random Forest Modelle konnte allerdings gezeigt werden, dass sie robust gegenüber Falschklassifikationen (zumindest bei geringem Anteil von Label Noise) in den Trainingsdaten sind (vgl. Pelletier et al., 2017: 20; Folleco et al., 2008: 3858). Das Konstruktionsprinzip von Random Forest Modellen aus Bootstrapping der Fälle (Bagging/Bootstrap Aggregation) und Randomisierung der untersuchten Variablen in den einzelnen Knoten der Entscheidungsbäume macht sie gegenüber Fehlern der Labels in den Trainingsdaten unempfindlicher als adaptive Boosting Verfahren, welche sich gerade auf Ausreißer bzw. Fehler bei der Klassifikation konzentrieren und versuchen diese zu minimieren. Es ist davon auszugehen, dass RS-Muster, welche durch die einfache Count-Methodik nicht erfasst wurden (und somit als falsch mit Non-RS gelabelt sind) auch unter ungünstigen Bedingungen nur einen kleinen Anteil der empirischen Daten bestimmen sollten. Daher sollte die Klassifikation anhand der Random Forest Modelle nicht übermäßig negativ beeinflusst werden.

Zusätzlich zu dieser rein empirischen Datenbasis wurden jeweils 10 000 Antwortmusterfälle simuliert, welche entweder dem ERS, ARS oder MRS Muster entsprechen (Simulationsmuster können aus Tabelle 1 entnommen werden).⁷

Diese simulierten Fälle wurden den Referenzdaten zugespielt und entsprechend ihrer jeweiligen RS-Musterstruktur klassifiziert. Bei dieser Simulation wurde der Anteil der von RS-Muster betroffenen Items variiert. Entweder waren alle $(\frac{4}{4})$, $\frac{3}{4}$, $\frac{2}{4}$ oder $\frac{1}{4}$ der 16 Frageitems mit

⁷Damit ist eine Balance zwischen der Anzahl der Fälle ohne RS und der Anzahl der Fälle mit RS-Mustern hergestellt.

Response Style Muster	Simulationsmuster	
	Rating Skala	Wahrscheinlichkeiten der Ratingkategorien
Extreme Response Style (ERS)	☒ □ □ □ ☒	[0.5 , 0.0 , 0.0 , 0.0 , 0.5]
Agreement Response Style (ARS)	☒ ☒ □ □ □	[0.7 , 0.3 , 0.0 , 0.0 , 0.0]
Midpoint Response Style (MRS)	□ □ ☒ □ □	[0.0 , 0.025, 0.95 , 0.025, 0.0]

Tabelle 1: Übersicht zu den simulierten Response Style Mustern

RS-Mustern generiert. Für die Simulationskonfigurationen in welchen nicht alle Items mit RS-Mustern generiert wurden, wurden die restlichen Frageitems durch ein Resampling aus den empirischen Daten des ESS 2002 hergestellt. Nach der Zusammenführung der empirischen und der simulierten Fälle in einen gemeinsamen Datensatz, wurde der Datensatz nach dem Zufallsprinzip in ein Trainingssample (80 %) und ein Validierungssample (20 %) geteilt um die ML-Modelle zu berechnen und anhand des Validierungssamples zu überprüfen.

4.3 Modellkonstruktion

Für die Random Forest Modelle wurde zunächst per Gridsearch (Brutforce-Berechnungen mit allen möglichen Ausprägungen) diejenigen Hyperparameter gesucht, welche die beste Voraussagegenauigkeit (bezüglich Out-of-Bag-Error und Validierungsstichprobenfehler) zeigten. Die Hyperparameter für Random Forest Modelle bestehen aus der Anzahl der verwendeten (zufällig gezogenen) Variablen in den Knoten des Baummodells und der Anzahl der zu berechnenden Bäume für das gesamte Ensemble Modell. Die Anzahl der Baummodelle wurde auf 500 festgelegt, wobei sich bereits bei deutlich weniger Bäumen die Fehlerraten stabilisierten (vgl. Bsp.-Abbildung 2).⁸ Für die Simulationskonfigurationen mit einem Anteil von $\frac{4}{4}$, $\frac{2}{4}$ oder $\frac{1}{4}$ an von RS betroffenen Items wurde mtry (die Anzahl der in jedem Baummodell zufallsgesteuert verwendeten Variablen) auf 4 und für die Konfiguration mit $\frac{3}{4}$ auf 3 festgelegt. Die Berechnungen wurden mit dem R Packet randomforest v4.6-14 durchgeführt.

5 Ergebnisse

Der hier vorgeschlagene Machine Learning basierte Ansatz kann dann als Alternative zu bisherigen Verfahren angesehen werden, wenn die Identifikation von Response Style Mustern zuverlässig genug gelingt und gleichzeitig inhaltlich-substantielle Antworten nicht falsch klas-

⁸Random Forest Modelle können nicht durch die Anzahl der Bäume overfitted werden. Es geht lediglich um eine ausreichend hohe Zahl für die Stabilisierung der Fehlerraten der Modelle.

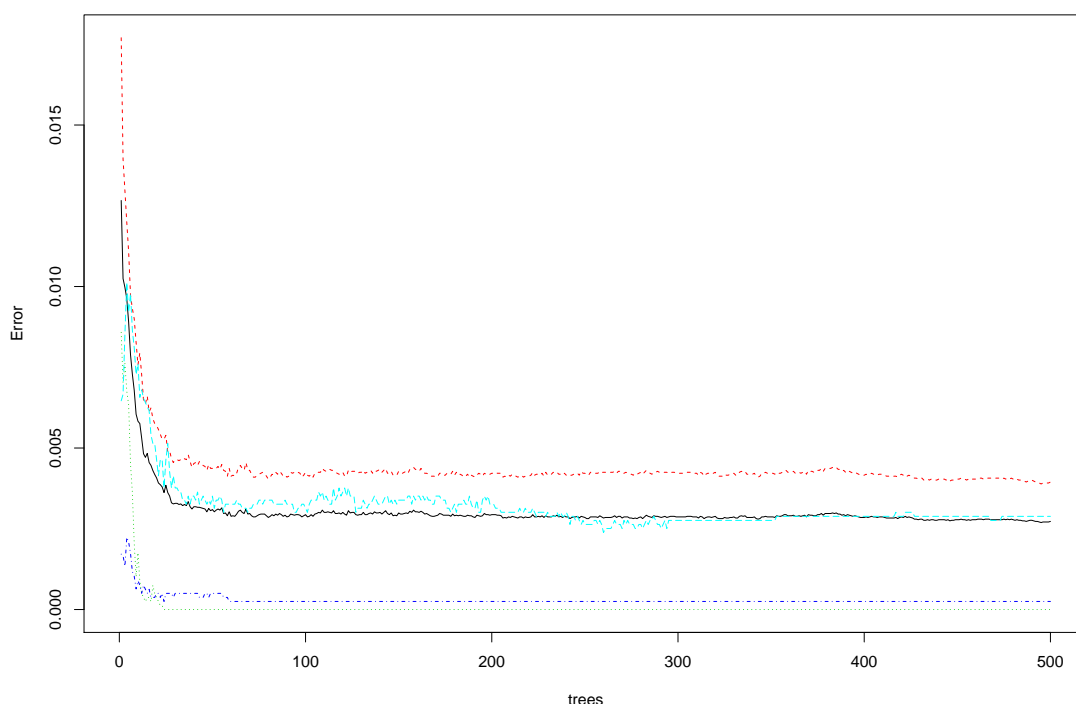


Abbildung 2: Beispielhafte Abbildung der Stabilisierung der Klassen-Fehlerraten in Abhängigkeit von der Anzahl der Baummodelle (Konfiguration $\frac{4}{4}$ RS-Anteil)

sifiziert werden. Diese Kriterien können anhand der Fehlerrate und anhand der sogenannten Confusion Matrix des Klassifikationsverfahrens abgelesen werden. Daher werden im Folgenden die Leistungsfähigkeit des Random Forest Modells bzgl. dieser Größen berichtet.

Die Klassifikationsperformanz der Random Forest Modelle fällt insgesamt äußerst gut aus. Für die Konfigurationen in welchen $\frac{4}{4}$ und $\frac{3}{4}$ der Items von Response Styles betroffen sind, bewegen sich alle klassenspezifische Fehlerraten deutlich unter einem Prozent. D. h. weder eine nennenswerte Fehlzuordnung von Non-RS-Fällen in die RS-Kategorien (oder andersrum) noch eine Fehlzuordnung zwischen den einzelnen Fallklassen der drei verschiedenen Formen von Response Styles lässt sich finden (vgl. hierzu Abbildung 3). Erst für die Konfiguration, in welcher lediglich $\frac{2}{4}$ der Items von RS-Muster betroffen sind, steigt die Fehlerrate langsam an. Allerdings bewegen sich die Fehlerraten auch hier lediglich im Wertebereich von 2.5 bis 0.1 %, mit der Kategorie MRS an der Spitze. Für die Konfiguration mit einem RS-Musteranteil von $\frac{1}{4}$ der Frageitems können schließlich Fehlerraten von maximal 6 % (für MRS) über 3.5 % (für Non-RS u. ERS) bis hin zu 1.6 % beobachtet werden. In der übergreifenden Gesamtbetrachtung über alle vier RS-Klassen hinweg, beträgt der durchschnittliche Klassifikationsfehler für die Modelle der Konfiguration $\frac{4}{4}$ bis $\frac{1}{4}$ jeweils 0.25, 0.31, 0.99 und 3.5 %.

Dieses Ergebnismuster der Fehlerraten aus Abbildung 3 lässt sich auch anhand der Con-

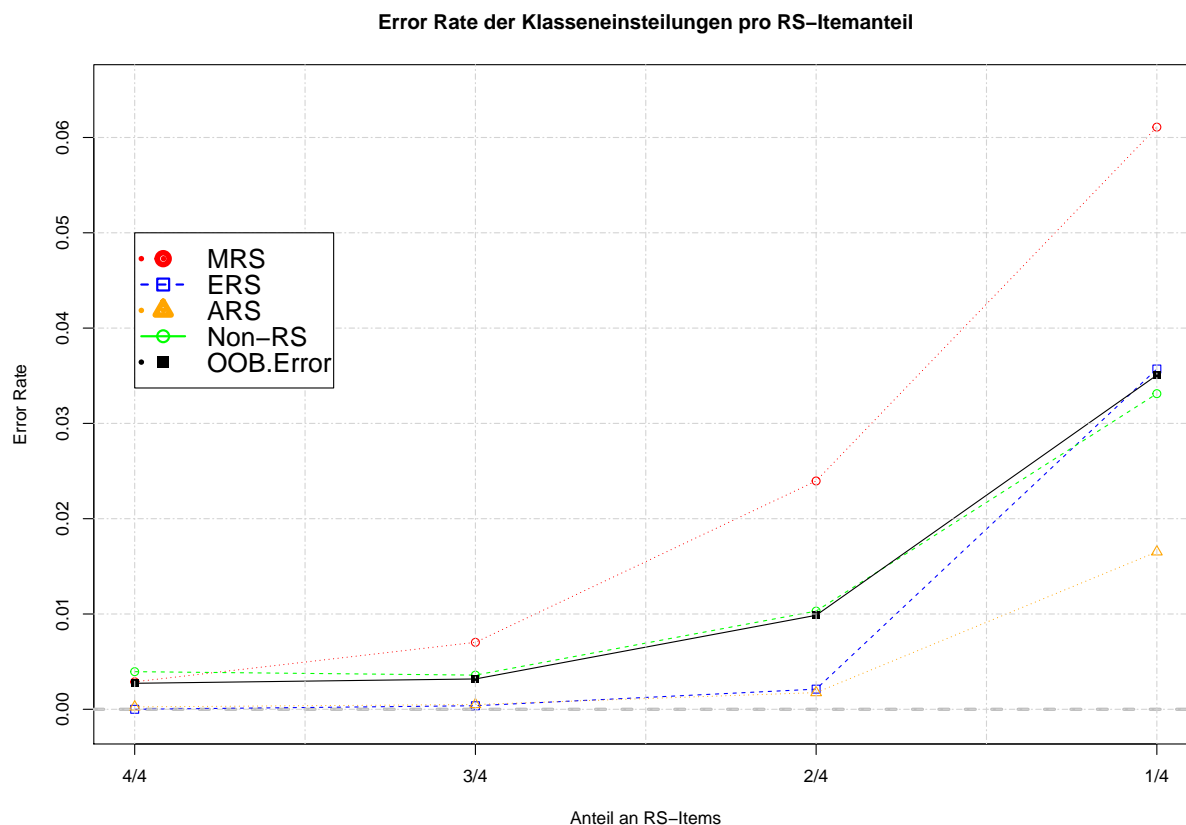


Abbildung 3: Error-Raten für die RS-Konfigurationsabstufungen. Eigene Darstellung.

fusion Matrix der einzelnen Konfigurationen ablesen (vgl. Abbildung 4). So lässt sich für die Konfigurationen, welche auf allen und nur auf $\frac{3}{4}$ der Frageitems von RS betroffen ist, kaum eine nennenswerte Anzahl an falsch zugeordneten Fällen finden. Erst für die $\frac{2}{4}$ und $\frac{1}{4}$ Konfiguration erreicht die absolute Anzahl der falsch klassifizierten Fälle einen erwähnenswerten Umfang. Diese fehlerhaften Klassifikationen konzentrieren sich auf falsch eingeteilte Non-RS Fälle, welche als RS-Muster geschätzt werden, und einige MRS-Fälle, welche als Non-RS klassifiziert wurden. Dies ist für die falsch klassifizierten MRS-Fälle insofern plausibel, als dass die synthetisierten MRS-Muster am wenigsten streng definiert wurden und auch geringfügige Abweichungen auf die Nachbarkategorien der Ratingskala in der Simulation berücksichtigt wurden (vgl. hierzu auch Tabelle 1). Auf diese ggf. zu liberale Definition scheinen die Random Forest Modelle insbesondere bei einem geringen Anteil von RS-Items sensitiv zu reagieren. Die einfachste Interpretation für falsch klassifizierten Non-RS-Fälle wäre eine geringere Modellleistung für diese Konfigurationen. Allerdings liegt auch eine alternative Deutung nahe. Hierzu muss nochmals auf das Problem des «Label Noise» in den empirischen Fällen eingegangen werden. So konnten für die empirischen Daten des ESS 2002 zwar durch die Count-Methode extreme RS-Fälle auf den meisten Items ausgeschlossen werden (vgl. Unter-

abschnitt 4.2), aber wenn sich RS-Muster nur auf einen kleinen Teil der Items zeigt, konnte dies nicht erkannt werden. Trotz dessen wurden empirischen Fälle, nach der Reduktion um extreme Fälle, als Non-RS klassifiziert. So besteht nun die Möglichkeit, dass die vermeintlich falsch klassifizierten Non-RS-Fälle tatsächlich doch einem RS-Muster folgen, weswegen sie von den Random Forest Modellen entsprechend klassifiziert wurden. Damit wäre die Fehlklassifikation kein wirklicher Fehler bzw. ein Hinweis auf geringe Modellperformanz, sondern ein Zugewinn durch das ML-Verfahrens. Diese Möglichkeit wurde untersucht, indem die Variablenverteilungen der falsch klassifizierten Non-RS-Fälle mit den Variablenverteilungsmustern der simulierten Fälle der entsprechenden RS-Kategorie verglichen wurden. Tatsächlich zeigt sich, dass sich die Verteilungen der vermeintlich falsch klassifizierten Fälle und ihr simulierter Gegenpart hinsichtlich der prozentualen Anteile über die Variablen sich in vielen Aspekten der Verteilung sehr gleichen (vgl. hierzu Abbildung 5 im Anhang für Simkonf. 1/4). Diese Ähnlichkeiten stechen besonders hervor, wenn man zusätzlich den Vergleich mit der Verteilungsform der Gesamtheit der empirischen Daten des ESS anstellt (vgl. Abbildung 1). Sicher kann dies nicht als ein formaler Test verstanden werden, aber die alternative Deutung im Sinne von Label-Noise wird hierdurch zumindest gestützt. Inwiefern es sich hierbei um mangelnde oder gar um besonders hohe Modelperformanz handelt, kann allerdings nur im Rahmen einer vollständigen Monte-Carlo-Studie beantwortet werden, welche Label-Noise durch eine gänzliche Synthetisierung der Datenbasis als alternative Erklärung ausschließen kann.

	1/4	2/4	3/4	4/4
Cut 0.75	295 (0.9 %)	232 (0.7 %)	144 (0.4 %)	109 (0.3 %)
Cut 0.50	604 (1.8 %)	360 (1.1 %)	226 (0.7 %)	245 (0.7 %)
Cut 0.25	1448 (4.5 %)	681 (2.1 %)	501 (1.5 %)	484 (1.5 %)
Vorausgesagte RS-Fälle in den empirischen Daten in Abhängigkeit der WHS-Schwellenwerte (N = 32 347)				

Tabelle 2: Random Forest Vorhersage auf empirische Daten

Diese trainierten ML-Modelle können im nächsten Schritt auf die ursprünglichen und vollständigen empirischen Daten angewendet werden, um fallspezifische Wahrscheinlichkeiten für das Vorhandensein von Response Style Antwortmustern zu bestimmen. Die Anwendung des ML-Modells auf die vollständigen Daten des ESS 2002 prognostiziert in Abhängigkeit vom

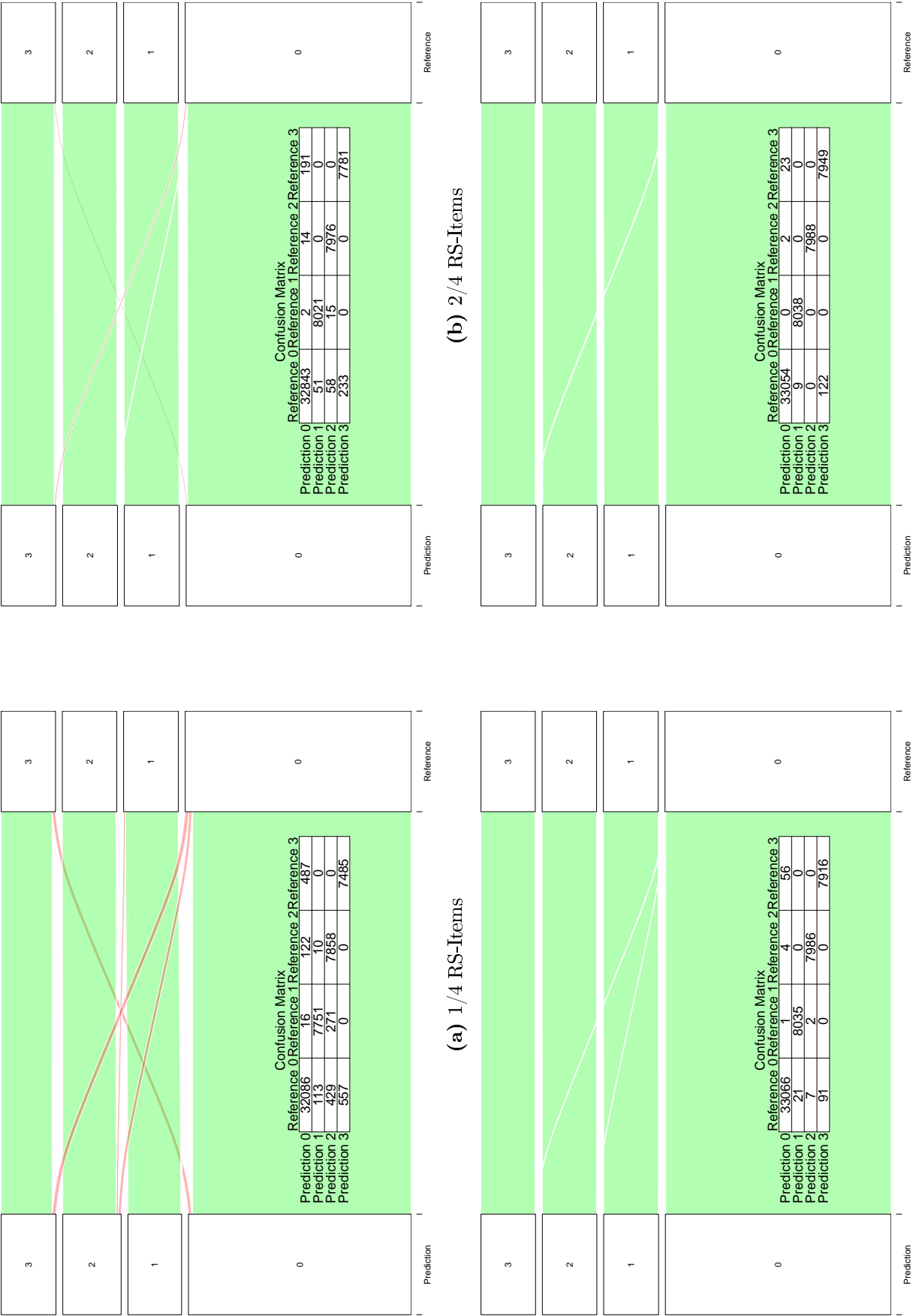


Abbildung 4: Plot-Visualisierung der Confusion Matrizen der Random Forest Modelle.

0 = Non-RS, 1 = ERS, 2 = ARS, 3 = MRS.

Korrelation mit Count Method		
	WHS aus RF	majority vote RF
4/4	0.65***	0.581***
3/4	0.62***	0.580***
2/4	0.53***	0.506***
1/4	0.38***	0.396***
Korrelationen zwischen RF u. Count Method nach WHS für RS u. Majority Vote des RS. *** $p < 0.01$; (N = 32 347)		

Tabelle 3: Kriteriumsvalidität bzgl. Count Method

angesetzten Grenzwert für die Wahrscheinlichkeit für Response Style Muster und des Umfangs der betroffenen Items Werte zwischen 0.3 bis 4.5 % (vgl. Tabelle 2). Für die mathematisch stringentste Variante der ML-Vorhersage (schlichter majority vote) werden folglich je nach Umfang des RS-Musters 1.5 bis 4.5 % der Fälle als Response Style Muster eingestuft (ohne Unterscheidung nach dem Typ des RS-Musters).

Da kein Zugriff auf den tatsächlichen Data Generating Process (DGP) der vollständigen empirischen Daten möglich ist, kann auch kein festes und absolutes Kriterium für Validität der Machine Learning basierten RS-Identifikation gefunden werden. Um dennoch eine Form von Validierung der durch das ML-Modell prognostizierten RS-Einstufung durchzuführen, wurde die Einstufung nach der einfachen Count-Methode als eine Form von Kriteriumsvalidität herangezogen. Die Korrelation zwischen der ML-Prognose und der Count-Klassifikation nimmt für Konfiguration mit $\frac{4}{4}$ RS-Items erwartungsgemäß den höchsten Wert an und sinkt für niedrigere RS-Itemanteile stufenweise ab (vgl. Tabelle 3). Da die Count Methode nur Fälle als RS-Muster klassifiziert, welche auf allen Items entsprechende Muster ausweisen, ist diese Zusammenhangsstruktur plausibel. Insgesamt spricht die Höhe und Differenz der Assoziationen für die Validität der ML-Prognose, auch wenn diese Form der Validierung nur als Approximation verstanden werden kann.

6 Diskussion, Ausblick und Fazit

Welcher praktische Nutzen kann nun auf Grundlage der Voraussageergebnisse der Random Forest Modelle gewonnen werden? Das Endergebnis des Trainings des ML-Verfahrens und der Vorhersage anhand dieses Modells ist eine Wahrscheinlichkeit der Klassenzugehörigkeit. Für jeden Fall werden dabei – auf Grundlage des ML-Modells – Wahrscheinlichkeiten für die Zugehörigkeit zu jeder untersuchten RS-Fallklasse geschätzt.

Diese Wahrscheinlichkeiten können wiederum genutzt werden um Fälle ab einem gewissen Schwellenwert auszuschließen um Verzerrungen von Mittelwerten, Varianzen und Assoziationen zu verringern. Alternativ (und weniger radikal) können diese Wahrscheinlichkeiten auch genutzt werden, um die empirischen Fälle für die Analyse zu gewichten und dadurch Verzerrungen, in Abhängigkeit von der Sicherheit über entsprechende Muster, zu verringern. Darüber hinaus kann die Wahrscheinlichkeit für RS-Muster auch als Kontrollvariable im Rahmen eines Analysesmodells mit spezifiziert werden, um Verzerrungen durch Auspartialisierung zu vermeiden oder die Stärke des verzerrenden Effektes hinsichtlich einer Zielgröße abzuschätzen. Andersherum können die Wahrscheinlichkeiten für RS-Muster auch als abhängige Größe untersucht werden, um mögliche Prädiktoren des entsprechenden Antwortverhaltens abzuschätzen und ggf. bei Folgeuntersuchungen zu kontrollieren oder entsprechende Gegenmaßnahmen anzuleiten. Neben der hohen Klassifikationsperformanz sind die wesentlichen Vorteile dieses Verfahrens zur Abschätzung der Wahrscheinlichkeiten, dass sie

- unabhängig vom spezifizierten Analysemodell sind,
- mehrere Formen von RS gleichzeitig berücksichtigen,
- keine zusätzlichen Frageitems benötigen,
- weitestgehend frei von expliziten Modellannahmen über die Genese von Response Style Mustern sind und,
- dass sie keine weitere Modellkomplexität erzeugen, welche die Schätzung und Konvergenz von Auswertungsmodellen negativ beeinflussen könnten.

Letztendlich ist die vorliegenden Arbeit allerdings nur eine Pilotstudie. Es hat gegolten zu erörtern, ob eine ML-basierte Identifikation von Response Styles möglich ist und wie erfolgversprechend die Performanz der Klassifikation anhand eines empirischen Datensatzes ausfällt. Diese ersten Ergebnisse bzgl. des Anwendungsbeispiels des ESS 2002 sind vielversprechend und rechtfertigen eine weitere Untersuchung. Darüber hinaus darf aber nicht vergessen werden, dass es sich lediglich um ein einzelnes Anwendungsbeispiel handelt. Die ideosynkratischen Besonderheiten der Datenbasis können hierbei eine maßgebliche Rolle spielen und nicht das Verfahren an und für sich. Darüber hinaus fehlt für eine umfängliche Bewertung des Verfahrens ein hartes Kriterium für die Validität des RS-Status der Fälle. Eine systematische Bewertung des Verfahrens kann allerdings auch nur anhand von einer vollständig syntetisierten Datenbasis einer Monte-Carlo-Untersuchung oder anhand einer entsprechend gestalteten

Survey-Experimentalstudie geschehen. Daher ist eine Exploration über das untersuchte Datenbeispiel hinaus noch verfrüht.

Weitere Untersuchungen sollten folglich zum Ziel haben, die Validität derartig gewonnener Klassifikation und die Universalität der Klassifikationsperformanz unter verschiedenen Datensituationen abzusichern. Das Mittel der Wahl hierfür wäre eine Monte-Carlo-Simulation, welche nicht nur eindeutige RS-Muster generieren könnte, sondern auch die Verteilungs- und Zusammenhangsstruktur der Non-RS-Fälle systematisch variiert. Damit könnte die Leistungsfähigkeit dieser Form von RS-Klassifikation verallgemeinert werden. Ein weiterer vielversprechender Ansatz wären Survey-Experimente, in welchen RS-Antwortmuster und inhaltliche Antworten gezielt randomisiert den Experimentaleinheiten zugeordnet werden. Dies könnte bspw. in Form von spezifischen Befragtenanweisungen und alternativen Maßwerten für RS-Muster geschehen. Damit könnte die externe Validität der Antwortmustersverteilung abgesichert werden und gleichzeitig alternative Kriterien der ML-Klassifikation, im Sinne von Paradata, wie Reaktionszeitmessung (etc.), genutzt und auf mögliche Voraussageverbesserung untersucht werden.

7 Literatur

- Baumgartner, H. & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2):143–156.
- Billiet, J. B. & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4):608–628.
- Birnbaum, B. (2012). *Algorithmic Approaches to Detecting Interviewer Fabrication in Surveys*. PhD thesis, University of Washington.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, 41(3):1–58.
- Eid, M. & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, 16(1):20–30.
- Falk, C. F. & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological methods*, 21(3):328–347.
- Folleco, A., Khoshgoftaar, T. M., van Hulse, J., & Bullard, L. (2008). Software quality modeling: The impact of class noise on the random forest classifier. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 3853–3859.
- Ghani, R. & Schierholz, M. (2017). Machine learning. In Ian Foster, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, & Julia Lane, editors, *Big data and social science: a practical guide to methods and tools*.
- Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by spanish-english bilinguals. *Cross-Cultural Research*, 33(4):369–381.
- He, J., van de Vijver, JR, F., Espinosa, A. D., & Mui, P. H. C. (2014). Toward a unification of acquiescent, extreme, and midpoint response styles. *International Journal of Cross Cultural Management*, 14(3):306–322.
- Jandura, O. (2018). Fake data? In Rössler, P. & Rossmann, C., editors, *Kumulierte Evidenzen*, pages 207–223. Springer Fachmedien Wiesbaden, Wiesbaden.
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles. *Journal of Cross-Cultural Psychology*, 36(2):264–277.
- Krause, T. (2019). *Multiple Indicator Growth Mixture Models: eine statistische Simulation*

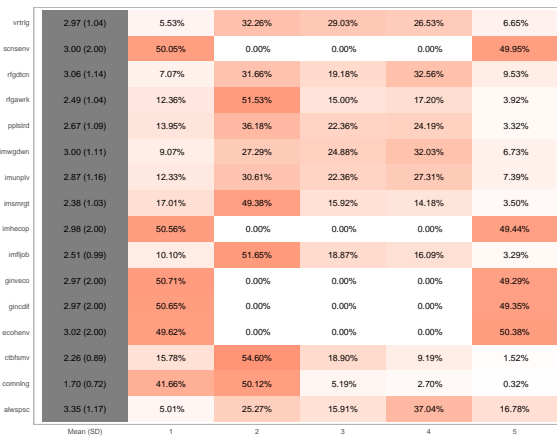
- zur Performanzevaluation für sozialwissenschaftliche Analysen. Dissertation, University of Stuttgart.
- Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44(7):1539–1550.
- Meiser, T. & Machunsky, M. (2008). The personal structure of personal need for structure. *European Journal of Psychological Assessment*, 24(1):27–34.
- Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, 22(1):93–119.
- Paulhus, D. L. (1990). Measurement and control of response bias. In Robinson, J. P., editor, *Measures of personality and social psychological attitudes*, Measures of social psychological attitudes. Academic Press, San Diego.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C., & Dedieu, G. (2017). Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing*, 9(2):173.
- Plieninger, H. (2017). Mountain or molehill? a simulation study on the impact of response styles. *Educational and psychological measurement*, 77(1):32–53.
- Plieninger, H. & Meiser, T. (2014). Validity of multiprocess irt models for separating content and response styles. *Educational and psychological measurement*, 74(5):875–899.
- Rammstedt, B., Danner, D., & Bosnjak, M. (2017). Acquiescence response styles: A multilevel model explaining individual-level and country-level differences. *Personality and Individual Differences*, 107:190–194.
- Rorer, L. G. (1965). The great response-style myth. *Psychological bulletin*, 63(3):129.
- Saris, W. E. & Aalberts, C. (2003). Different explanations for correlated disturbance terms in mtmm studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(2):193–213.
- Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach to evaluating the quality of measurement instruments: The split-ballot mtmm design. *Sociological Methodology*, 34(1):311–347.
- Schimmack, U., Böckenholt, U., & Reisenzein, R. (2002). Response styles in affect ratings: making a mountain out of a molehill. *Journal of personality assessment*, 78(3):461–483.
- van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales. *Journal of Cross-Cultural Psychology*, 35(3):346–360.
- van Vaerenbergh, Y. & Thomas, T. D. (2013). Response styles in survey research: A literature

- review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2):195–217.
- Weijters, B. (2006). *Response styles in consumer research*. PhD thesis, Ghent University.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3):236–247.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological methods*, 15(1):96–110.
- Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, 36(3):409–422.
- Wetzel, E. & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, 33(5):352–364.
- Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*, 47(2):178–189.
- Wickrama, K. K., Lee, T. K., O’Neal, C. W., & Lorenz, F. O. (2016). *Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide*. Multivariate Applications Series. Taylor and Francis, s.l.

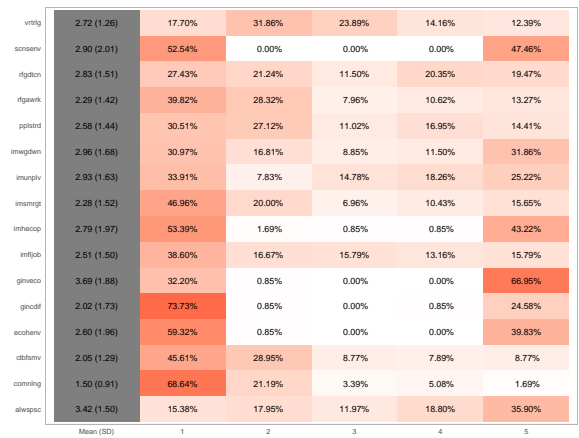
A Anhang

Variable	Item label
ginveco	The less that government intervenes in the economy, the better it is for [country]
gincdif	The government should take measures to reduce differences in income levels
sensenv	Modern science can be relied on to solve our environmental problems
ecohenv	Economic growth always ends up harming the environment
imhecop	People who come to live and work here generally harm the economic prospects of the poor more than the rich
imfljob	People who come to live and work here help to fill jobs where there are shortages of workers
imunply	If people who have come to live and work here are unemployed for a long period, they should be made to leave
imsmrgrt	People who have come to live here should be given the same rights as everyone else
imwgdwn	Average wages and salaries are generally brought down by people coming to live and work here
ctbfsmv	All countries benefit if people can move to countries where their skills are most needed
pplstrd	It is better for a country if almost everyone shares the same customs and traditions
vrtrlg	It is better for a country if there are a variety of different religions
comnlng	It is better for a country if almost everyone is able to speak at least one common language
alwspsc	Communities of people who have come to live here should be allowed to educate their children in their own separate schools if they wish
rfgawrk	While their applications for refugee status are being considered, people should be allowed to work in [country]
rfgdten	While their cases are being considered, applicants should be kept in detention centres

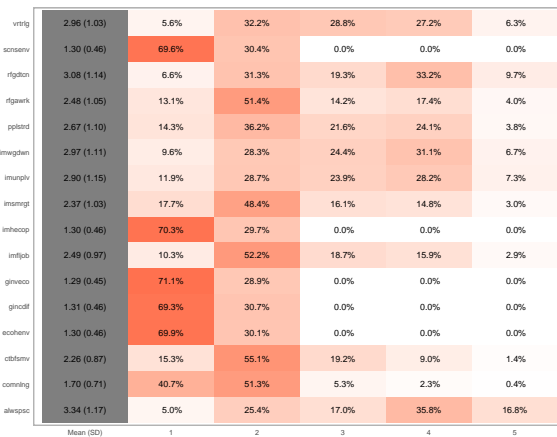
Tabelle 4: Übersicht zu den Frageitems des ESS 2002



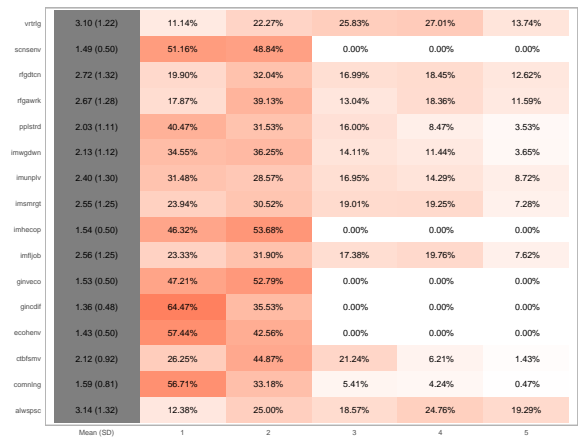
(a) simulierte Faele ERS



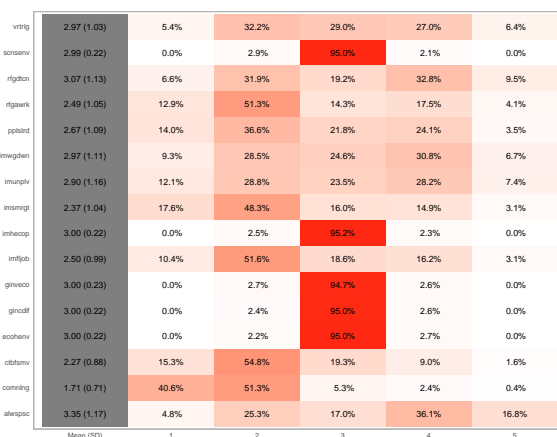
(b) Missklassifizierte Faele ERS



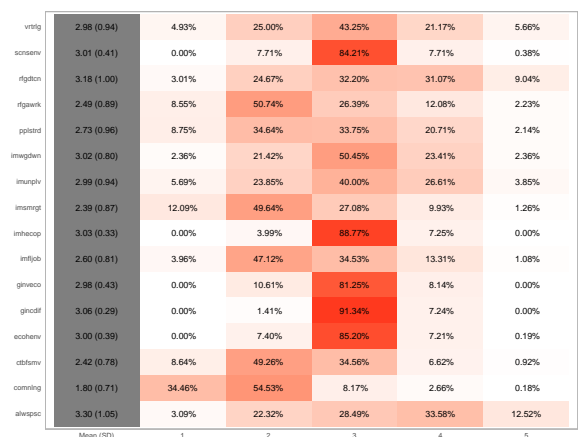
(c) simulierte Faele ARS



(d) Missklassifizierte Faele ARS



(e) simulierte Faele Simkonf MRS



(f) Missklassifizierte Faele MRS

Abbildung 5: Vergleich zwischen simulierten und missklassifizierten RS-Faellen (1/4 Konfig.)

SISS:**Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart**

<https://www.sowi.uni-stuttgart.de/forschung/schriftenreihe-des-instituts-fuer-sozialwissenschaften-siss/>

- No. 47/2020 Machine Learning basierte Response Style Identifikation:
eine simulations-statistische Pilotstudie.
Thomas Krause
- No. 46/2019 Informationskriterien zur latenten Klassenidentifikation für Multiple Indicator
Growth Mixture Models.
Thomas Krause
- No. 45/2017 Die Angst vor dem Fremden. Ein Vignettenexperiment zur Bedrohungswahrnehmung.
Ergebnisse eines studentischen Projektseminars.
Alexander Geisler, Alexandra Möck, Rebecca Nell, Svenja Wiechmann, Thomas Krause
- No. 44/2016 Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder
Nicht-Imputation? Eine Anleitung für die Regressionspraxis mit SPSS.
Dieter Urban, Jochen Mayerl, Andreas Wahl
- No. 43/2014 Meta-Analyse mit Mehrebenenmodellen unter Verwendung von HLM.
Joachim Fiebig, Dieter Urban
- No. 42/2013 Panelanalyse mit Mehrebenenmodellen. Eine anwendungsorientierte Einführung.
Thomas Krause, Dieter Urban
- No. 41/2012 Meta-Analyse und "publication bias". Eine empirisch-statistische Fallstudie.
Joachim Fiebig, Gerrit Rüdebusch, Dieter Urban
- No. 40/2011 Pädosexuelle Viktimisierung und pädosexuelle Straffälligkeit. Ein „harter“ empirischer
Hypothesentest.
Dieter Urban, Joachim Fiebig
- No. 39/2010 Binär-logistische Regressionsanalyse.
Grundlagen und Anwendung für Sozialwissenschaftler.
Jochen Mayerl, Dieter Urban
- No. 38/2010 Das Working-Poor-Problem in Deutschland.
Empirische Analysen zu den Ursachen von Armut trotz Erwerbstätigkeit.
Leonie Hellmuth, Dieter Urban
- No. 37/2010 Der Bystander-Effekt in alltäglichen Hilfesituationen: Ein nicht-reaktives Feldexperiment.
Katrin Alle, Jochen Mayerl
- No. 36/2008 Berufseintritt und Berufssituation von Soziologieabsolventen der Universität Stuttgart.
Deskriptive Ergebnisse einer Absolventenbefragung aus dem Jahr 2007.
Jochen Mayerl, Dieter Urban
- No. 35/2007 Antwortlatenzzeiten in TRA-Modellen. Zur statistischen Erklärung von (Geld)-Spendenver-
halten.
Dieter Urban, Jochen Mayerl
- No. 34/2006 Zur soziologischen Erklärung individuellen Geldspendens.
Eine Anwendung und Erweiterung der Theory of Reasoned Action unter Verwendung von
Antwortlatenzzeiten in einem Mediator-Moderator-Design.
Jochen Mayerl
- No. 33/2006 Ist Bildung gleich Bildung? Der Einfluss von Schulbildung auf ausländerablehnende
Einstellungen in verschiedenen Alterskohorten.
Klaus Hadwiger

SISS:**Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart**

<https://www.sowi.uni-stuttgart.de/forschung/schriftenreihe-des-instituts-fuer-sozialwissenschaften-siss/>

- No. 32/2005 Analyzing cognitive processes in CATI-Surveys with response latencies: An empirical evaluation of the consequences of using different baseline speed measures.
Jochen Mayerl, Piet Sellke, Dieter Urban
- No. 31/2005 Persönliche Netzwerke in der Systemtheorie.
Jan A. Fuhse
- No. 30/2004 Erhöht ein Opfer-Täter-Zyklus das Risiko, Sexualstraftaten als pädosexuelle Straftaten zu begehen? Ergebnisse einer ereignisanalytischen Pilotstudie.
Dieter Urban, Heiko Lindhorst
- No. 29/2003 Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden? Eine empirische Analyse computergestützter Telefoninterviews.
Jochen Mayerl
- No. 28/2003 Systeme, Netzwerke, Identitäten. Die Konstitution sozialer Grenzziehungen am Beispiel amerikanischer Straßengangs.
Jan A. Fuhse
- No. 27/2002 Probleme bei der Messung individueller Veränderungsraten.
13 empirisch und methodisch induzierte Effekte, die es schwierig machen, Veränderungen von generalisierten Bewertungen zu ermitteln.
Dieter Urban
- No. 26/2002 Vertrauen und Risikoakzeptanz. Zur Relevanz von Vertrauen bei der Bewertung neuer Technologien.
Martin Slaby, Dieter Urban
- No. 25/2002 Risikoakzeptanz als individuelle Entscheidung. Zur Integration der Risikoanalyse in die nutzentheoretische Entscheidungs- und Einstellungsforschung.
Martin Slaby, Dieter Urban
- No. 24/2001 Ist Glück ein affektiver Sozialindikator subjektiven Wohlbefindens? Dimensionen des subjektiven Wohlbefindens und die Differenz zwischen Glück und Zufriedenheit.
Jochen Mayerl
- No. 23/2001 Religiosität und Profession. Longitudinale Analysen zur Entwicklung des religiösen Selbstbildes bei Erzieherinnen.
Heiko Lindhorst
- No. 22/2001 Differentielle Technikakzeptanz, oder: Nicht immer führt die Ablehnung einer Technik auch zur Ablehnung ihrer Anwendungen. Eine nutzentheoretische und modell-statistische Analyse.
Martin Slaby, Dieter Urban
- No. 21/2001 Unser „wir“ - ein systemtheoretisches Modell von Gruppenidentitäten.
Jan A. Fuhse
- No. 20/2000 Längsschnittanalysen mit latenten Wachstumskurvenmodellen in der politischen Sozialisationsforschung.
Dieter Urban

SISS:**Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart**

<https://www.sowi.uni-stuttgart.de/forschung/schriftenreihe-des-instituts-fuer-sozialwissenschaften-siss/>

- No. 19/2000 Die Zufriedenheit von Stuttgarter Studierenden mit ihrer Lebens- und Wohnsituation. Erste deskriptive Ergebnisse einer sozialwissenschaftlichen Studie zu allgemeinen und bereichsspezifischen Zufriedenheiten der Studierenden des Campus Vaihingen und des Campus Hohenheim.
Projektgruppe Campus: Slaby, M.; Grund, R.; Mayerl, J.; Noak, T.; Payk, B.; Sellke, P.; Urban, D.; Zudrell, I.
- No. 18/1999 Role Models and Trust in Socio-Political Institutions: A Case Study in Eastern Germany, 1992-96.
Joachim Singelmann, Toby A. Ten Ayck, Dieter Urban
- No. 17/1998 Zur Interaktion zwischen Befragten und Erhebungsinstrument. Eine Untersuchung zur Konstanz des Meinungsurteils von Befragten im Interviewverlauf.
Martin Slaby
- No. 16/1998 Technikeinstellungen: gibt es die überhaupt? Ergebnisse einer Längsschnittanalyse von Bewertungen der Gentechnik.
Dieter Urban
- No. 15/1998 Bewertende Einstellungen zur Gentechnik: ihre Form, ihre Inhalte und ihre Dynamik. Kurzbericht zu Ergebnissen des Forschungsprojektes „Einstellungen zur Gentechnik“.
Dieter Urban, Uwe Pfenning, Joachim Allhoff
- No. 14/1997 Politische Sozialisation im Transformationsprozeß: Die Entwicklung demokratiebezogener Einstellungen von ostdeutschen Jugendlichen und deren Eltern zwischen 1992 und 1996.
Barbara Schmidt, Dieter Urban, Joachim Singelmann
- No. 13/1997 Die soziale Vererbung von Ausländer“feindlichkeit“. Eine empirische Längsschnittanalyse der intra- und intergenerativen Transmission von sozialen Einstellungen.
Dieter Urban und Joachim Singelmann
- No. 12/1997 Gentechnik: „Fluch oder Segen“ versus „Fluch und Segen“. Bilanzierende und differenzierende Bewertungen der Gentechnik in der öffentlichen Meinung.
Dieter Urban und Uwe Pfenning
- No. 11/1996 Einstellungsmessung oder Einstellungsgenerierung? Die Bedeutung der informationellen Basis bei Befragten für die empirische Rekonstruktion von Einstellungen zu gentechnischen Anwendungen.
Martin Slaby
- No. 10/1996 Jugend und Politik im Transformationsprozeß. Eine Fallstudie zur Stabilität und Veränderung von politischen Einstellungen bei ostdeutschen Jugendlichen zwischen 1992 und 1995.
Dieter Urban, Joachim Singelmann, Helmut Schröder
- No. 09/1995 Social Indicators in a Nonmetropolitan County: Testing the Representativeness of a Regional Nonrandom Survey in Eastern Germany.
Dieter Urban, Joachim Singelmann
- No. 08/1995 Handbuch zur empirischen Erhebung von Einstellungen/Kognitionen zur Bio- und Gentechnologie (inklusive Diskette)
(zweite, überarbeitete und erweiterte Auflage)
Uwe Pfenning, Dieter Urban, Volker Weiss

SISS:**Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart**

<https://www.sowi.uni-stuttgart.de/forschung/schriftenreihe-des-instituts-fuer-sozialwissenschaften-siss/>

- No. 07/1995 Latent Attitude Structures Directing the Perception of New Technologies.
An Application of SEM-Methodology to the Construction of Attitude Measurement Models
Related to Technologies of Prenatal Genetic Engineering and Testing.
Dieter Urban
- No. 06/1995 Regionale Strukturprobleme.
Sammelband der Beiträge zum Symposium vom 13. und 14. Oktober 1994.
Frank C. Englmann (Hrsg.)
- No. 05/1995 Die Bedeutung Neuronaler Netze in der Ökonomie.
Hermann Schnabl
- No. 04/1994 Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht aus den neuen
Bundesländern.
Udo Kornblum
- No. 03/1994 Neue württembergische Rechtstatsachen zum Unternehmens- und
Gesellschaftsrecht.
Udo Kornblum
- No. 02/1994 Report on the German Machine Tool Industry.
Frank C. Englmann, Christian Heyd, Daniel Köstler, Peter Paustian
with the assistance of Susanne Baur and Peter Bergmann
- No. 01/1994 "Vertrauen" - soziologisch betrachtet. Ein Beitrag zur Analyse binärer
Interaktionssysteme.
Peter Antfang, Dieter Urban